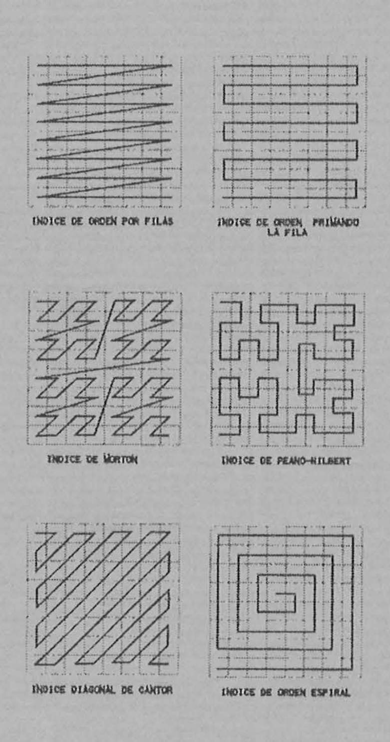
BASES DE DATOS Y GESTORES DE BASES DE DATOS PARA LOS SISTEMAS DE INFORMACIÓN GEOGRÁFICA

por

PILAR CHÍAS NAVARRO



CUADERNOS
DEL INSTITUTO
JUAN DE HERRERA
DE LA *ESCUELA DE*
ARQUITECTURA
DE MADRID

5-11-05

BASES DE DATOS Y GESTORES DE BASES DE DATOS PARA LOS SISTEMAS DE INFORMACIÓN GEOGRÁFICA

por

PILAR CHÍAS NAVARRO

CUADERNOS
DEL INSTITUTO
JUAN DE HERRERA
DE LA *ESCUELA DE*
ARQUITECTURA
DE MADRID

5-11-05

CUADERNOS
DEL INSTITUTO
JUAN DE HERRERA

0 VARIOS
1 ESTRUCTURAS
2 CONSTRUCCIÓN
3 FÍSICA Y MATEMÁTICAS
4 TEORÍA
5 GEOMETRÍA Y DIBUJO
6 PROYECTOS
7 URBANISMO
8 RESTAURACIÓN

NUEVA NUMERACIÓN

5 Área
11 Autor
05 Ordinal de cuaderno (del autor)

Bases de datos y gestores de bases de datos para los sistemas de información geográfica

Instituto Juan de Herrera.
Escuela Técnica Superior de Arquitectura de Madrid.
Gestión y portada: Nadezhda Vasileva Nicheva
CUADERNO 174.01 / 5-11-05
ISBN: 84-9728-117-9
Depósito Legal: M-44372-2004

BASES DE DATOS Y GESTORES DE BASES DE DATOS PARA LOS SISTEMAS DE INFORMACIÓN GEOGRÁFICA

Índice

1. Las bases de datos y los gestores de bases de datos 3

1.1 Las características y organización de una base de datos en un SIG 6

2. Las estructuras de los archivos de datos 12

2.1 La organización no ordenada 13

2.2 La estructura secuencial 14

2.3 Organización fragmentada 16

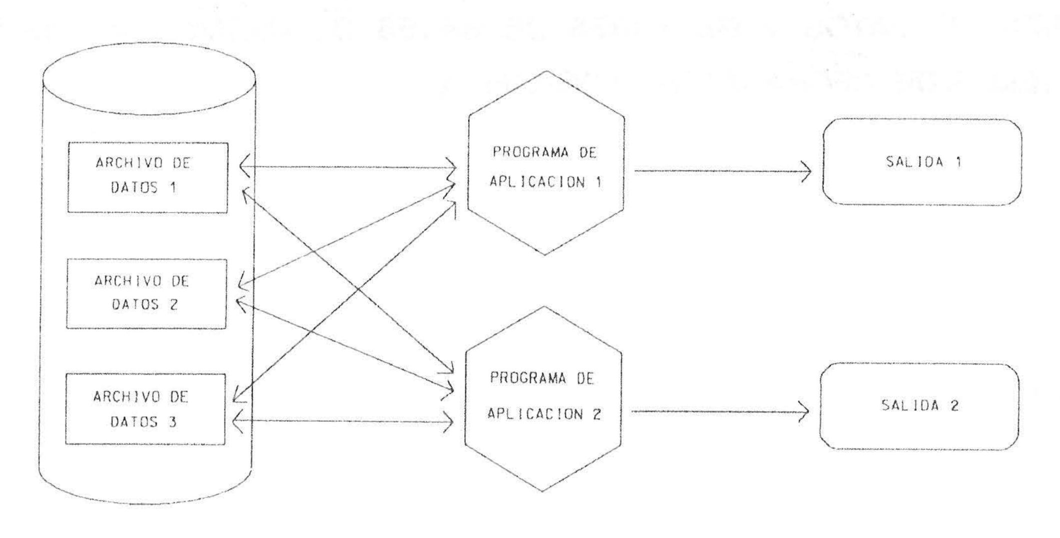
2.4 Organización indexada 16

3. Las relaciones existentes entre los datos: "uno a uno", "uno a muchos", "muchos a muchos" 19

4. Ordenaciones uni- y bidimensionales 20

Comentarios bibliográficos 28

Bibliografía citada 29



Archivo 1
Archivo 2
Archivo 3

Sistema de Gestión de la Base de datos DBMS (ODBC)

Programa de Aplicación 1

Programa de Aplicación 2

Salida 1

Salida 2

Prof. Chías

29

BASES DE DATOS Y GESTORES DE BASES DE DATOS

Conocidas las posibilidades de definir matemáticamente las entidades y relaciones más frecuentes dentro del dominio de aplicación o ámbito espacial de estudio, es necesario diseñar el *modelo de datos* que se va a utilizar.
Éste define el "almacén" de los datos (base de datos) a todos los niveles del SIG. Es decir, que el modelo de datos actúa tanto al nivel del diseño concreto de la base de datos (*modelo lógico de datos*), como al de su aplicación sobre un sistema y plataforma particulares y con unos dispositivos de almacenamiento determinados (*modelo físico*).

Por tanto el modelo de datos tiene que poder ser procesable por el ordenador, a la vez que ha de permitir al usuario interactuar con él y acceder a la base de datos para efectuar las extracciones y las consultas oportunas. Y todo ello ha de realizarse con la mayor rapidez y seguridad.

Precisamente porque el proceso ha de ser informatizable, resulta conveniente recordar previamente algunos conceptos generales relativos a estos almacenes de datos y a sus *gestores*, para después avanzar en las diferentes posibilidades de definición del modelo de datos de acuerdo con las geometrías anteriormente expuestas y con las necesidades concretas que cada SIG plantea.

1. Las bases de datos y los gestores de bases de datos.

Una base de datos es un almacén de datos que puede ser compartido por varios usuarios. Los datos pueden ser números, cadenas de caracteres, textos, imágenes, configuraciones espaciales e incluso sonidos; es decir, todo aquéllo que pueda ser medido, grabado, y a lo que se le pueda asignar un significado.

Los datos que poseen un significado en virtud de su contexto y de sus relaciones con otros grupos de datos, se han transformado en información.



Todos los datos de un SIG, tanto si son espaciales como si son atributivos, se almacenan en bases de datos; para ello se unifican siguiendo criterios lógicos y se estructuran de determinadas maneras. Asimismo los datos pueden estar físicamente distribuidos en lugares diferentes, así como ser accesibles por diferentes usuarios.

Los primeros *Sistemas de Base de Datos* (DBS) fueron programas que se desarrollaron para realizar un peque□o conjunto de funciones bien definidas y utilizando un grupo de datos determinados. En general, se limitaban a transformar el primer grupo de datos en otro a través de un largo y complicado proceso, por lo que en definitiva eran como "calculadoras gigantes".

Ello planteaba serios problemas porque:

- El programa accedía directamente a la base de datos, lo que requería conocer su estructura de almacenamiento.
- Existía redundancia en las instrucciones de acceso a la base, porque tenían que repetirse.
- Y al carecer de un control centralizado y existir la posibilidad de un acceso simultáneo por parte de varios usuarios, se favorecía la capacidad de alteración y degradación de los datos.

A la vista de los inconvenientes, que redundaban en falta de eficacia, surgió una segunda tendencia que consideraba el ordenador como un gran almacén de datos, y para ello se dotó a aquél de ciertas capacidades para facilitar el almacenamiento, la recuperación, la modificación y el análisis.
Con este fin se diseñaron los *Sistemas de Gestión de las Bases de Datos* (*Database Management System,* DBMS), que son programas de propósito general que crean y mantienen las bases de datos actuando como un control central sobre todas las interacciones que tienen lugar entre la base de datos y los programas de aplicación.

Las características de los gestores de bases de datos son:
- El control centralizado de las condiciones de accesibilidad a los datos, que impide la acción de ciertas funciones para mantener la calidad de los mismos. Ello no excluye la posibilidad de que varios usuarios compartan simultáneamente los datos, pero siempre de un modo controlado. En general soportan una gran variedad de usos, que van desde las consultas esporádicas e imprevisibles, hasta las habituales y expertas.
- Facilitan el acceso al usuario porque poseen unos mecanismos de interacción entre éste y los datos (*interfaces*), que facilitan las operaciones. En general, los usuarios pueden interactuar con la base de datos por medio de un lenguaje interactivo de consultas (como por ejemplo el SQL, *Structured or Standard Query Language*), o a través de un programa de aplicación en el que se hayan integrado uno o más modos de llamar a la base de datos (*menús*). Las interacciones suelen consistir en introducir, modificar o eliminar datos, y recuperarlos.
- Permiten definir los *modelos de datos.*
- Disminuyen la redundancia de órdenes.

- Proporcionan diferentes formatos de presentación de los datos, de acuerdo con las necesidades concretas de los usuarios, y aunque compartan los mismos datos.
- Facilitan el dise□o de nuevos programas de aplicación que utilizan las bases de datos.

La "arquitectura" genérica de este tipo de bases de datos multiusuario, que permite trabajar a distintos niveles y compartir los datos desde distintos puntos de vista, se denomina ANSI-SPARC (del *American National Standards Insitute*).

1.1 Características y organización de una base de datos en un SIG.

Para que una base de datos resulte operativa, debe ser:

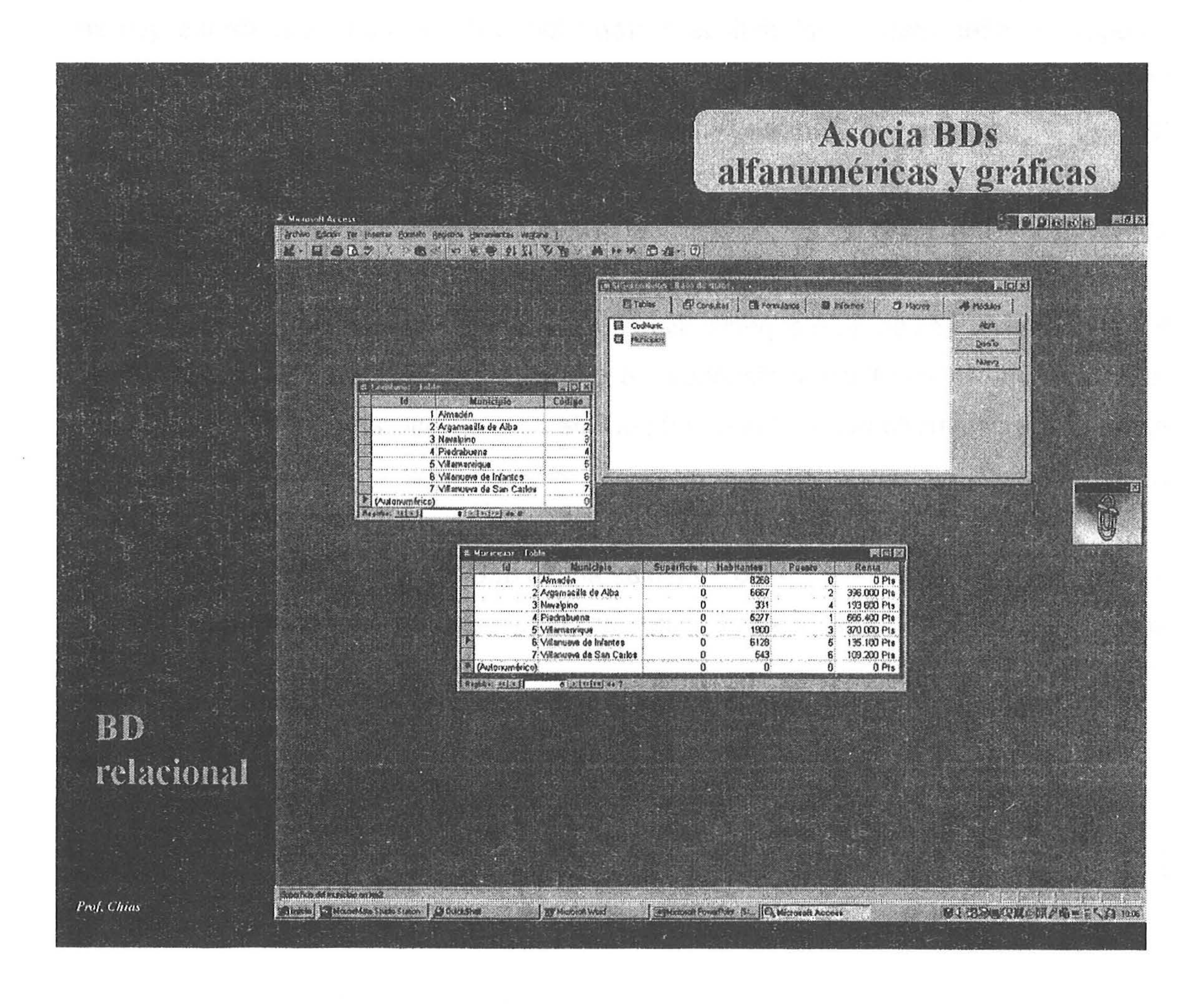
- *Segura*: previniendo accesos no autorizados y permitiendo diferentes grados de accesibilidad. Por ejemplo, un usuario de un SIG sobre la Ciudad Universitaria debería poder leer sólo los datos de carácter general y los relativos a su centro concreto, pero no debe poder modificarlos ni leer los correspondientes a otros centros si ello no se desea.
- *Disponible*: debe poder ser accesible cuando se necesite, y estar a salvo de imprevistos como un corte de energía.
- *Correcta y consistente*: los datos que se incorporen al sistema han de ser ciertos en sí mismos y en relación con los restantes datos. Este aspecto puede constituir un verdadero problema cuando se utilizan datos antiguos que pueden llegar a ser contradictorios con otros más actuales.

  Por ejemplo, si el SIG de la Ciudad Universitaria utilizara datos sobre la edificación existente hace una década, definiría como solar sin edificación la parcela que ocupan hoy las viviendas de profesores del Paseo de las Moreras, lo que proporcionaría información errónea sobre la superficie edificada, o contradictoria si se mezclase con la información referente a la edificabilidad consumida. Por ello, cuanto más se sepa de los datos que se manejan, más fácil será detectar posibles errores.

Por otra parte, cada capa debe ser lógicamente consistente; es decir, que dentro de cada capa han de estar correctamente definidas las relaciones topológicas existentes entre los elementos o los sucesos espaciales. Por ejemplo, cada polígono que represente a una parcela deberá ser cerrado y poseer un único identificador que sirva de referencia para asociarle los atributos.

Análogamente, la consistencia lógica debe existir también entre las diferentes capas de información: por ejemplo, dadas la capa de parcelación y la de propietarios del suelo, el trazado de los polígonos que representan las parcelas debe ser coincidente en ambas.

- Y *a prueba de tecnologías*: los datos no deben verse afectados por los cambios tecnológicos ni de los componentes físicos del sistema, ni de los programas.



Por otra parte, la organización típica de una base de datos consiste en una colección de *archivos* o *ficheros* con denominación propia, que el ordenador es capaz de crear, manipular y eliminar.
Un archivo es una secuencia de *registros*, generalmente del mismo tipo: por ejemplo, "parcelas" en el ámbito de la Ciudad Universitaria.

Los *registros* son secuencias de datos correspondientes a una misma entidad lógica. Por ejemplo, el archivo "Parcelas - Estado Actual" se compone de tantos registros (P1, P2,..Pn) como parcelas haya diferenciadas en el momento en el dominio de aplicación. Si todos los registros poseen igual tamaño en bytes, el archivo se compone de registros *de longitud fija.*
A su vez, cada registro incorpora datos relativos a una serie de *campos.* Éstos recogen las diferentes características o propiedades de las entidades de las que se tienen datos. Por ejemplo, de cada parcela se pueden obtener datos relativos al "uso", "superficie", "propietario", "edificabilidad", etc.

Un campo cuya longitud varíe de acuerdo con el contenido, se denomina campo de *longitud variable.* Es el caso habitual de los que contienen comentarios.
Entre los campos existe uno especial, llamado el *campo clave*, que es único para cada registro y que lo identifica y distingue de entre los demás sin ambigüedad. En el ejemplo que desarrollamos podría un número de parcela, utilizando números naturales no repetidos.

Todo ello se representa por medio de *tablas* en las que las filas son los registros, y las columnas son los campos. Y son las casillas de la tabla las que se rellenan con los datos.

	CAMPO				
REGISTRO					

Siguiendo con el ejemplo, la estructura del archivo "Parcelas" sería:

ARCHIVO PARCELAS DE LA CIUDAD UNIVERSITARIA

	USO	SUPERFICIE	PROPIETARIO	EDIFICABILIDAD
P1				
P2				
....				
Pn				

Rellenándose posteriormente con los datos las casillas en blanco de la tabla.

También resulta frecuente codificar ciertos datos de modo que se pueda ahorrar espacio de almacenamiento.
Por ejemplo, si asignamos al uso "docente" el valor 100, al de "investigación" el valor 200, etc., en la tabla aparecerán tres dígitos en lugar de las cadenas de caracteres más largas, y se habrá creado un campo de longitud fija que favorece la compacidad en el almacenamiento de los datos.

La unidad más pequeña de información que se graba en un disco se definió anteriormente como un "bloque" (*disk block*). Los archivos se ubican físicamente en un disco asignándoseles *disk blocks* para contener los registros. Si el disk block es más pequeño que el registro, éste tendrá que extenderse a lo largo de varios bloques, pero es más frecuente que cada bloque contenga varios registros.
Si varios registros sucesivos se sitúan en bloques adyacentes, se dice que sus posiciones son contiguas (*contiguous allocation*); pero esta organización puede alterarse cuando se mueven los registros, por lo que se recurre a otros tipos de colocación como la *vinculada* (*linked allocation*), que relaciona ciertos bloques entre sí.

El tiempo que el ordenador emplea en transferir un *block* desde o hacia un disco es importante, sobre todo cuando se manejan grandes cantidades de datos, y se puede descomponer en tres componentes:

- El *tiempo de búsqueda*, que es el que emplea el movimiento mecánico de la cabeza lectora hasta encontrar la pista correcta, y que es el factor que más influye en la búsqueda de un dato desde un dispositivo de almacenamiento secundario.
- El *tiempo de latencia*, que es el tiempo que tarda el disco en girar hasta la posición adecuada.
- Y el *tiempo de transferencia a la CPU*, que es el que se requiere para introducir el block en la CPU.

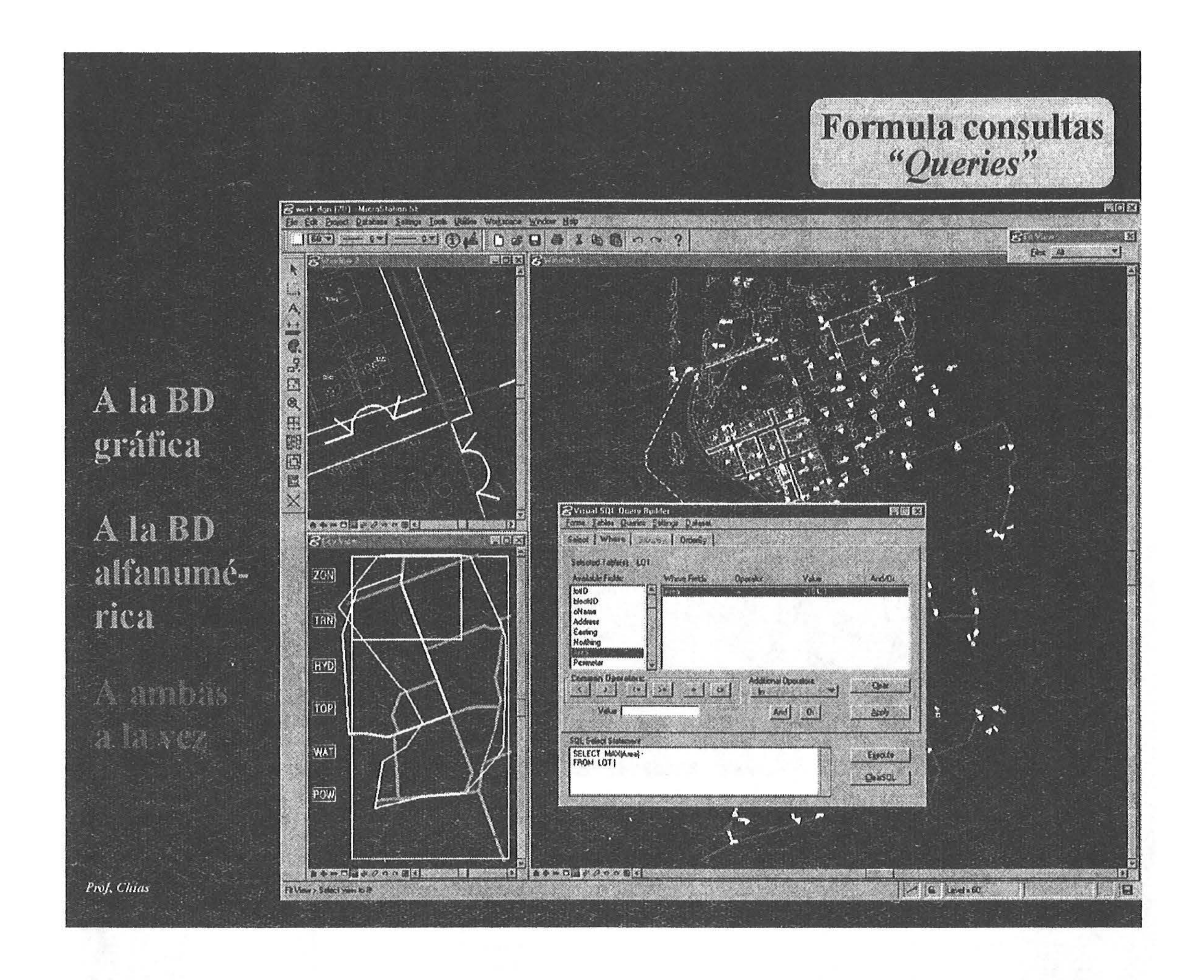
Por tanto, para reducir la influencia del tiempo de búsqueda las BDs se diseñan de modo que en lo posible los datos que tengan que acceder juntos a la CPU estén próximos en el disco, o bien se les dota de algún tipo de índice que remita la búsqueda directamente a los archivos que se requieren.

En general, cuando se habla de la *organización de los archivos* no sólo se trata de la organización física de los registros en el disco -o unidad de almacenamiento secundario-, de cómo se hayan introducido los datos en él, sino también en el modo en que están vinculados los bloques de registros.

Una vez situados los archivos en el dispositivo de almacenamiento, se pueden buscar por medio de los llamados *métodos de acceso*, que determinan el modo en el que el usuario va a manipular los archivos, y que consisten en un conjunto de operaciones primarias muy conocidas como son "abrir" (*open*), "encontrar" (*find*), "leer" (*read*), "borrar" (*delete*), "modificar" (*modify*), "insertar" (*insert*) y "cerrar" (*close*).

Un método de acceso complejo se compondrá de una secuencia de tales operaciones. Por ejemplo, para cambiar la "superficie" de una parcela cuyo identificador se conoce -por ejemplo, P81-, habrá que *abrir* el archivo "Parcelas", *buscar* el registro de la parcela P81, *modificar* el dato correspondiente al campo "superficie", y *cerrar* el archivo.

Por tanto, las bases de datos se diseñan generalmente para ser flexibles y para poder responder a una gran variedad de consultas. En los SIG es frecuente que algunas de estas consultas no sean previsibles de antemano, por lo que deben proporcionar una base estadística potente para hacer frente a los futuros patrones de búsqueda posibles.



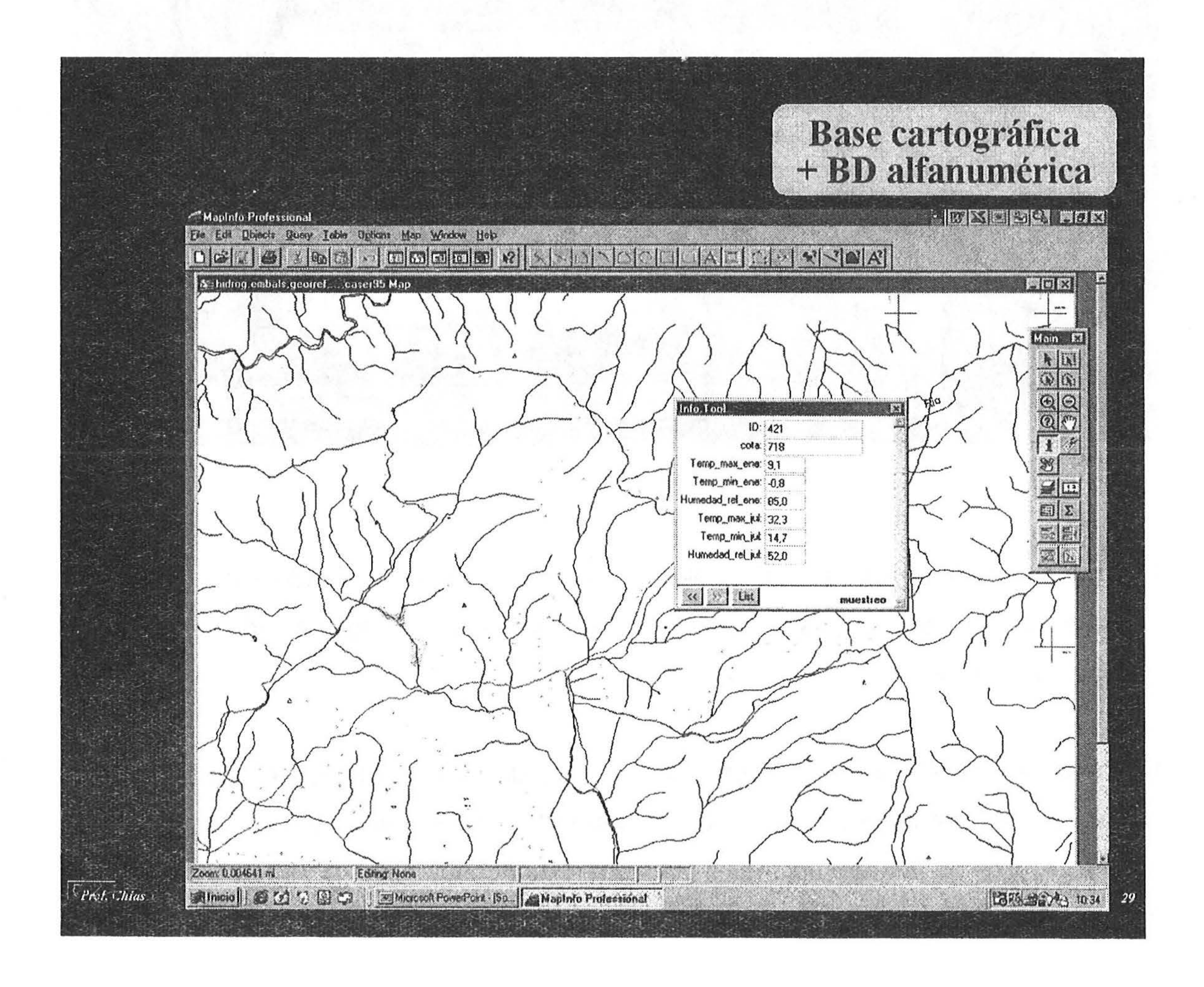
Las bases de datos de los SIG suelen ser muy dinámicas, variando sus contenidos con frecuencia, lo que resta eficacia a los métodos de organización por simple grabación sucesiva.

De acuerdo con el movimiento previsible de los datos en el SIG, existen diferentes maneras de estructurar los datos dentro de ellos, que tienen importantes consecuencias en las operaciones introducción de nuevos datos y de recuperación de los ya existentes.

2. Las estructuras de los archivos de datos: no ordenadas, secuenciales, fragmentadas e indexadas

Las entidades o fenómenos que se estudian y que se hallan dentro del dominio de aplicación, poseen dos componentes claramente diferenciables cuales son la componente *espacial* y la *atributiva*.

El SIG integra ambas componentes como datos dentro de archivos.



El modo de introducir los datos proporciona una estructura a los archivos que resulta importante a la hora de recuperar aquéllos. Son cuatro los modos habituales de estructurar los archivos: el modo *no ordenado*, el modo *ordenado o secuencial*, el modo *fragmentado* y el *indexado*.

2.1 *La organización no ordenada.*

Una organización de archivos no ordenada supone que los nuevos registros que se insertan en el archivo lo hacen en la posición siguiente del disco, tanto a partir del último block utilizado, como a partir de uno nuevo. Es una *lista simple* y la forma más sencilla de almacenamiento, porque cada nuevo dato sólo se a□ade a la lista de datos ya existente, resultando por ello bastante eficaz a la hora de introducir los datos.

El problema surge al buscarlos, pues al no existir ningún tipo de estructura aparte del orden de entrada, hay que comenzar la búsqueda desde el principio del archivo, en un *acceso secuencial.* Es decir, que para una lista de *n* registros, se requiere una media de $(n + 1)/2$ operaciones de búsqueda hasta encontrar el registro que se desea.
Por ejemplo, para los edificios de la Universitaria podemos construir la siguiente tabla con una estructura no secuencial.

ESTRUCTURA NO ORDENADA

Nº REGISTRO	Nº EDIFICIO	USO	CONSERVACIÓN	CATÁLOGO (PROTECCIÓN)
3	3	100	Regular	Singular
1	2	100	Buena	Integral
4	1	800	Mala	Estructural II
5	4	200	Buena	Ambiental
2	5	300	Regular	Estructural I

Esta estructura resulta aceptable cuando sólo se requieren listados; por ejemplo, si se quisiera imprimir una ficha de cada parcela.

Los problemas surgen cuando se quieren efectuar accesos directos y aleatorios, ya que se requiere información específica de los registros que se buscan.

Por ejemplo, si se quieren localizar las parcelas que tienen un nivel de protección dado en el Catálogo de la Edificación -supongamos que "Estructural I"-, es necesario examinar los valores que adopta el campo correspondiente en cada uno de los registros hasta que aparezca el que se desea.
Este tipo de búsqueda es *lineal* y debe evitarse en lo posible porque consume mucho tiempo. De hecho, si tuviéramos que buscar una parcela determinada entre un total de 100 registros -uno por parcela-, habría que realizar una media de (100 + 1)/2 operaciones de búsqueda.

Además, es un tipo de estructura que origina vacíos cuando se borran registros, y aunque se pueden volver a llenar, el proceso de inserción es muy lento.

2.2 *La estructura secuencial.*

En cambio, en la *estructura secuencial u ordenada* al menos uno de los campos está ordenado según una secuencia, que puede ser incluso la de introducción de los datos, pero que puede ser también alfabética, etc.
Siguiendo el ejemplo de los edificios, la codificación de tres dígitos asignada al campo de los usos permite ordenar los registros siguiendo la secuencia 100, 200,... 900. Este campo "ordenador" tendrá que poder ordenarse completamente -según los números enteros, según el orden lexicográfico de las cadenas de caracteres, etc.

La ventaja de este sistema es que permite búsquedas binarias aplicando un algoritmmo de búsqueda binario. Es además más rápida -el número de accesos se puede calcular con la función logarítmica $\log_2 (n)$ cuando no son mayoría los registros que verifican la condición de búsqueda-, porque generalmente comienza por la mitad del archivo e inmediatamente se sabe si ha de continuar en sentido ascendente o descendente.
Por ejemplo, una condición de búsqueda podría ser el de los edificios cuyo uso es "igual" al "300" -equivalente al definido como "administrativo universitario". Si contamos con un total de 1000 blocks en los que se ha almacenado el archivo de los

usos de las parcelas, una búsqueda lineal requeriría una media de 500 acccesos a los blocks, mientras que si los registros estuviesen ordenados por el uso, se requerirían aproximadamente $\log_2 1000 \tilde{}$ 10 accesos.

Muchas búsquedas se basan en condiciones de intervalo, en lugar de en condiciones de igualdad. Por ejemplo, se pueden localizar todas las parcelas cuyas superficies estén comprendidas entre los 2000 y los 4999 m^2. Para ello el campo ordenador tiene que ser el de "superficies".
Obviamente, si se realizase la búsqueda por otro campo, ésta volvería a ser lineal.

El problema principal aparece cuando hay que introducir nuevos datos, pues si no se les ha previsto un sitio dentro del orden, se requiere una reordenación; lo mismo sucede con los registros borrados, cuyo espacio puede ser llenado de nuevo.

En el ejemplo de los edificios del Campus, la estructura de la tabla será la siguiente:

ESTRUCTURA SECUENCIAL

USO	Nº REGISTRO	Nº EDIFICIO	CONSERVACIÓN	CATÁLOGO (PROTECCIÓN)
100	1	2	Buena	Integral
100	3	3	Regular	Singular
200	5	4	Buena	Ambiental
300	2	5	Regular	Estructural I
800	4	1	Mala	Estructural II

2.3 *Organización fragmentada.*

Un *archivo fragmentado* se organiza utilizando una función fragmentadora (*hash fuction*) que transforma los valores del campo fragmentado en una dirección dentro de un block en el cual se coloca el registro.

Por ejemplo, supongamos que contamos con 1000 blocks para el archivo de parcelas. Se puede fragmentar el campo del identificador de las parcelas -supongamos que es un número entero de 4 dígitos- tomando los dos últimos para que actúen como etiqueta del block: la parcela 2035 se situará en el block etiquetado como 35.

Para buscar los registros en un campo fragmentado sólo se necesitará, por tanto, un único acceso al disk block.

Esta técnica es muy sencilla, pero se complica cuando los blocks se saturan, y resulta difícil asegurar una distribución adecuada si no hay espacio en todos ellos. Por ello se necesita saber de antemano el número de blocks que se van a necesitar antes de elegir la función de fragmentación -por ejemplo, tomando el último dígito, o los tres últimos.

2.4 *Organización indexada.*

Las organizaciones en lista y secuencial ordenada, requieren que los datos se recuperen de acuerdo con un atributo o campo-clave. Pero en muchas aplicaciones, y entre ellas las vinculadas con la información geográfica, las entidades o fenómenos individuales no sólo tienen un campo o atributo identificador -un nombre o una clave numérica-, sino también otros atributos asociados que además, son frecuentemente los que se quieren buscar.

Por ejemplo, supongamos que tenemos las parcelas de la Universitaria ordenadas por el tipo de uso que pueden soportar, pero que necesitamos información sobre el nivel de Protección asociado: salvo que se disponga otra estrategia de búsqueda, se tendrá que proceder por el método de la lista.

Los índices (*indexes*) aceleran la actuación sobre los otros campos, porque actúan igual que los índices de los libros, ordenando los nombres de los objetos indexados de modo que se puedan localizar rápidamente.
El archivo suplementado con un índice se denomina *archivo de datos* (*data file*), y en él el índice actúa como un campo independiente llamado campo indexador (*indexing field*).
También se crea un *campo indicador* (*pointer field*) que contiene las direcciones de los disk blocks que contienen el valor del índice. Además, un índice puede actuar sobre más de un campo.

La estructura *indexada* se puede aplicar a cualquier archivo tomando como referencia un campo determinado, lo que facilita la búsqueda de los datos; además, si se introducen nuevos datos sólo hay que indexar de nuevo el archivo.

Es la estructura más eficaz y la que proporciona mayor libertad en el manejo del archivo, porque permite indexar tantas veces como sea necesario.

ESTRUCTURA INDEXADA

Índice

CAMPO INDEXADOR	CAMPO INDICADOR
P30 P31 P75	• • • ....

Archivo de datos

ID_Parcela	Nombre	Superficie (m^2)	Uso
0001	Investigación Veterinaria	200.502	400
0002	Investigaciones Agrarias	33.576	400
0003	Puerta de Hierro	78.146	900
....	....	....	....
0010	Junta de Energía Nuclear Este	171.318	500

BLOCK 1

0030	Prácticas Agrónomos Norte	76.657	100
0031	Estadística	46.941	100
....	....	....	...
0045	Libre Almudena Sur	6.562	900

BLOCK 2

....

0075	Colegio Pío XII	8.664	700
....	....	....	....
0100	Colegio Casa del Brasil	10.592	700

BLOCK 100

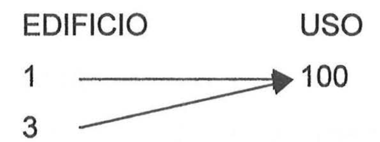
3 Las relaciones existentes entre los datos: "uno a uno", "uno a muchos", "muchos a muchos"

Si se observan con detenimiento los datos introducidos, se puede deducir la existencia de cierto tipo de relaciones entre los diferentes campos.
Análogamente tales relaciones pueden reproducirse entre los campos y las entidades, y se pueden agrupar en los tipos siguientes:

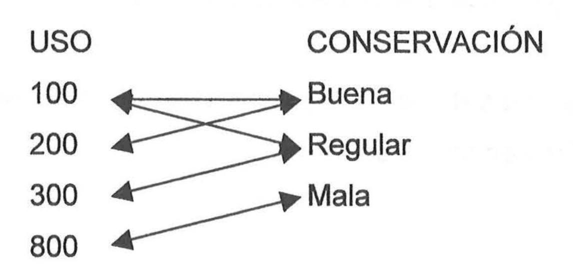
- La *relación "uno a uno"* es biunívoca, de modo que a cada valor de un campo corresponde únicamente un valor en el otro campo. Un ejemplo de ello es la relación que existe entre cada entidad y un único identificador:

IDENTIFICADOR ⟺ PARCELA

- La *relación "uno a muchos"* permite que a diferentes valores de un mismo campo les pueda corresponder el mismo valor en otro campo. Por ejemplo, varias parcelas pueden tener el mismo uso:



- La *relación "muchos a muchos"* es aquélla en la que cualquier valor de un campo se corresponde con varios de otro campo, y viceversa. Tal sería la relación existente entre los campos "uso" y "conservación":



- Finalmente, la *relación de dependencia funcional* es aquélla que se produce cuando para conocer el valor de uno o varios campos es necesario conocer previamente los valores de otro u otros campos. Por ejemplo, para conocer el porcentaje de ocupación de las parcelas se requiere un campo previo que recoja su área, y otro que posea los valores de la superficie ocupada por las edificaciones en cada parcela.

El conocimiento y definición de tales relaciones resulta esencial a la hora de diseñar el modelo de datos concreto que interesa para una determinada aplicación.

4 Ordenaciones uni- y bidimensionales

Los comentarios anteriores son extensivos a cualquier base de datos de propósito general, en la que los archivos son multidimensionales, es decir, que cada entidad posee varios atributos independientes -aunque ya se ha puesto de manifiesto que hay casos en los que puede existir cierta dependencia, que puede eliminarse manipulando los archivos y manteniendo los de origen.
Los datos espaciales presentan una relación de dependencia entre ellos que se expresa en el espacio euclídeo por medio de la métrica euclídea.

Consideremos por ejemplo los datos de algunos puntos clave de la Universitaria que se representan de forma tabular y en forma de malla.

Id	Centro	Este	Norte
11	UNED Norte	750	2440
13	Agrícolas	1600	1200
15	Escuela de Telecomunicaciones	1970	2590
18	Arquitectura Técnica	1420	780
20	Veterinaria	570	2170
21	Geografía	1050	1780
23	Letras	1450	2230
25	Ciencias	1970	2250
28	CEU Arquitectura	2610	2580
31	Estadística	860	1790
32	Escuela de Caminos	1350	1790
35	Escuela de Montes	2370	2250
42	Ciencias de la Educación	2850	2400
55	Ciencias de la Información	1750	1630
70	Escuela de Agrónomos	1790	1190
72	Medicina	2050	1405
77	UNED Suroeste	1050	380
80	Bellas Artes	1190	830
81	Escuela de Arquitectura	1500	840
84	Escuela de Aeronáuticos	2180	980
85	Escuela de Navales	2220	790
91	UNED Sur	1220	180
113	Forestales	2230	2020
115	Ingenieros Técnicos Aeronáuticos	2230	980
117	Investigación Politécnica	835	600

Para buscar entre estos datos, podemos utilizar dos tipos de condiciones:

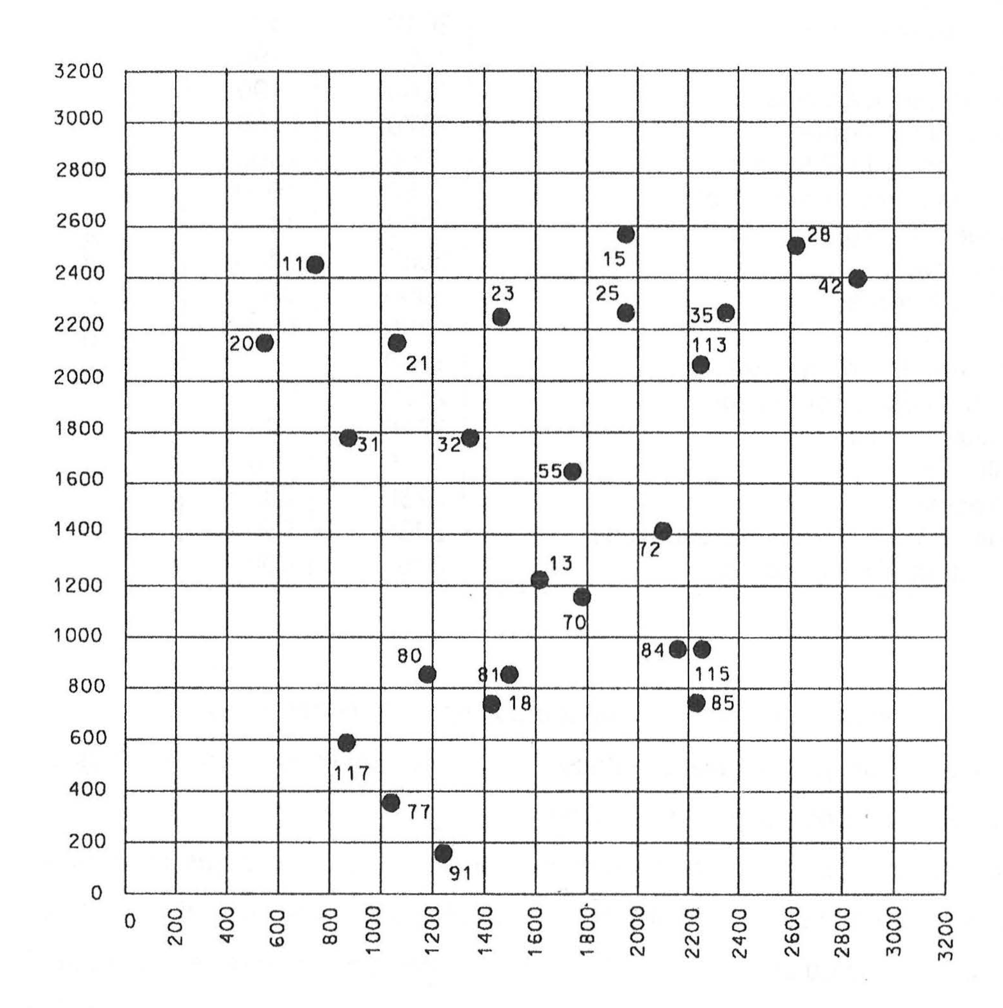
- *Condiciones de punto:* que satisfacen todos los registros cuyas referencias espaciales coincidan con el punto dado.
- *Condiciones de intervalo:* que verifican todos los registros cuyas referencias espaciales se encuentren dentro de un intervalo dado. Este intervalo puede ser una forma cualquiera, aunque se suele utilizar una superficie rectangular definida por las coordenadas de dos vértices opuestos, o un círculo definido por el centro y el radio.

Las consultas que se pueden hacer son:

1º *No espaciales*: "Localizar el punto correspondiente a la Escuela de Arquitectura".

Esta consulta proporciona una posición en el espacio, pero se contesta buscando en el archivo un campo no espacial (Escuela de Arquitectura ⇒ punto 81).

Para ello se puede utilizar cualquiera de los métodos de acceso comentados para las bases de datos de propósito general.



Localización de algunos puntos de interés (centros docentes universitarios) en la Ciudad Universitaria

2º *Consulta del punto*: "Buscar cualquier lugar situado en la posición (1500,840)". Esta consulta utiliza una condición espacial de búsqueda. Si no se establece ningún orden, aquélla será lineal:

Paso 1. Abrir el archivo de los Centros de la Ciudad Universitaria

Paso 2. Mientras haya registros que examinar, seguir los pasos 3.1 a 3.4

Paso 3.1 Acceder al siguiente registro

Paso 3.2 Si el valor del primer campo de coordenadas es 1500

Paso 3.3 Entonces Si el valor del segundo campo coordenado es 840

Paso 3.4 Entonces obtener del registro el nombre del centro

3º *Consulta de intervalo*: "Buscar cualquier lugar en el rectángulo cuyos vértices suroeste y noreste son, respectivamente, (1200,600) y (1600,1000)".

La secuencia de la consulta lineal será, en este caso:

Paso 1. Abrir el archivo de Centros de la Universitaria

Paso 2. Mientras haya registros que examinar, seguir los pasos 3.1 a 3.4

Paso 3.1 Acceder al siguiente registro

Paso 3.2 Si el valor del primer campo de coordenadas está dentro del intervalo [1200,1600]

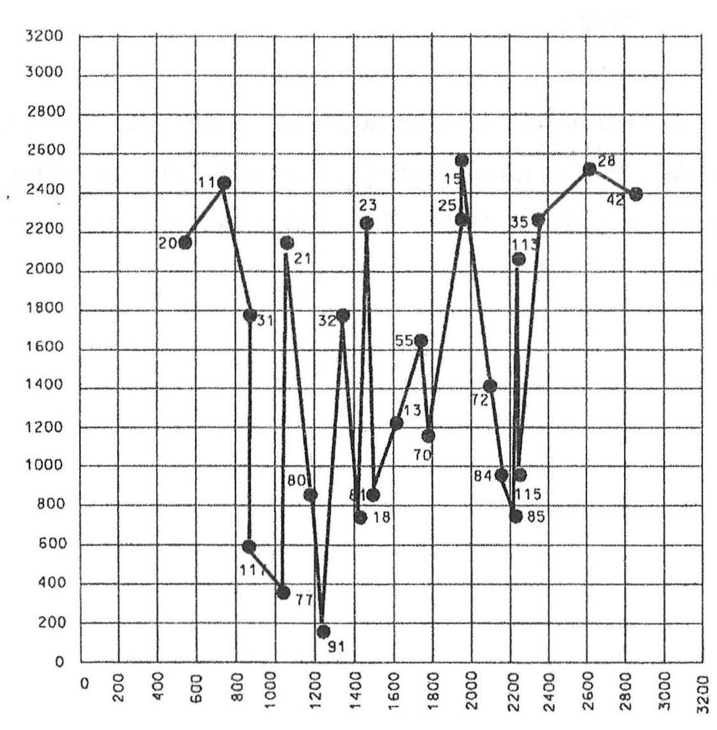
Paso 3.3 Entonces Si el valor del segundo campo de coordenadas está dentro del intervalo [600,1000]

Paso 3.4 Entonces obtener del registro el nombre del centro

Pero también se puede agilizar la búsqueda construyendo dos índices para los dos campos espaciales de coordenadas (índices Este y Norte); como cada uno afecta a un sólo campo, se pueden considerar unidimensionales, al igual que las operaciones de búsqueda que con ellos se realizan.

Índice Este

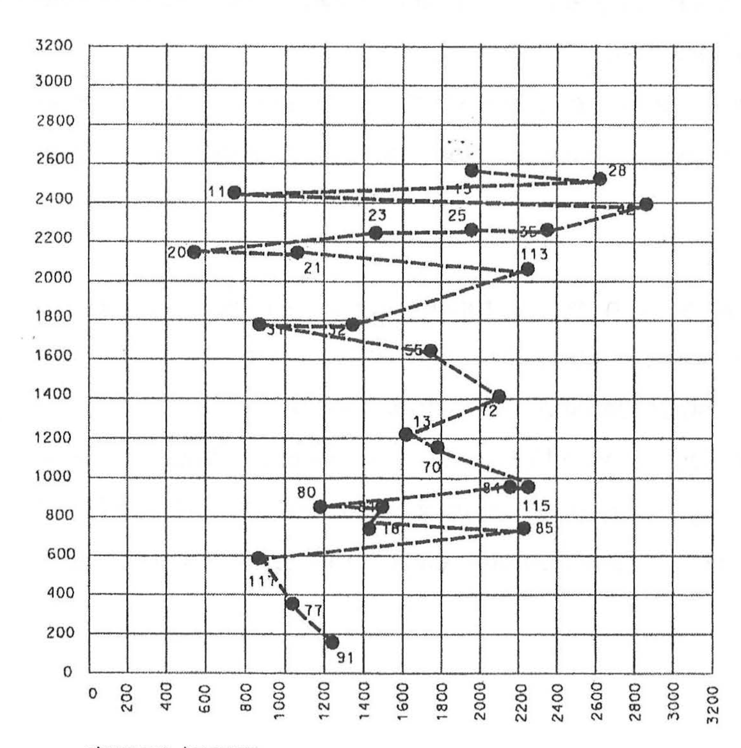
Centro	*Este*
Veterinaria	570
UNED Norte	750
Investigación Politécnica	835
Estadística	860
Geografía	1050
UNED Suroeste	1050
Bellas Artes	1190
UNED Sur	1220
Escuela de Caminos	1350
Arquitectura Técnica	1420
Letras	1450
Escuela de Arquitectura	1500
Agrícolas	1600
Ciencias de la Información	1750
Escuela de Agrónomos	1790
Escuela de Telecomunicaciones	1970
Ciencias	1970
Medicina	2050
Escuela de Aeronáuticos	2180
Escuela de Navales	2220
Forestales	2230
Ingenieros Técnicos Aeronáuticos	2230
Escuela de Montes	2370
CEU Arquitectura	2610
Ciencias de la Educación	2850



INDICE ESTE

Indice Norte

Centro	*Norte*
UNED Sur	180
UNED Suroeste	380
Investigación Politécnica	600
Arquitectura Técnica	780
Escuela de Navales	790
Bellas Artes	830
Escuela de Arquitectura	840
Escuela de Aeronáuticos	980
Ingenieros Técnicos Aeronáuticos	980
Escuela de Agrónomos	1190
Agrícolas	1200
Medicina	1405
Ciencias de la Información	1630
Geografía	1780
Estadística	1790
Escuela de Caminos	1790
Forestales	2020
Veterinaria	2170
Letras	2230
Ciencias	2250
Escuela de Montes	2250
Ciencias de la Educación	2400
UNED Norte	2440
CEU Arquitectura	2580
Escuela de Telecomunicaciones	2590



Las dos secuencias de puntos que resultan de los órdenes respectivos se muestran con diferentes tipos de línea en la ilustración adjunta.

Si volvemos a la 2ª Consulta, se dirige la búsqueda por el índice Este hasta que encuentre los registros cuyo valor es 1500; después accede a la tabla general hasta encontrar aquéllos cuyo valor coincide con 840.

En cuanto a la 3ª Consulta, se puede establecer un intervalo de búsqueda en el índice Este entre los valores 1200 y 1600, del que resultará una lista de puntos indicadores del archivo de datos.

Para cada uno, el sistema accederá al registro correspondiente y comparará el valor de la coordenada Norte, y extraerá aquéllas que se encuentren en el intervalo [600,1000].

Como se puede comprobar, en estas búsquedas sólo se ha utilizado un índice.

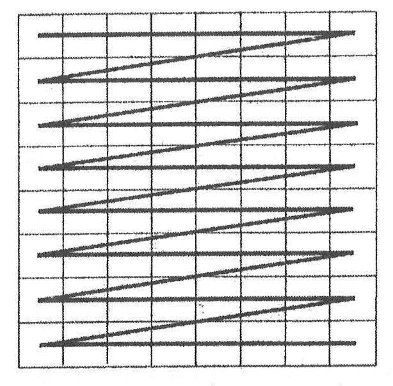
Por tanto, lo que se puede hacer es buscar otro tipo de índice que permita ordenar en las dos dimensiones.

El primer problema que se plantea es que el sistema de almacenamiento de los datos en el sistema es básicamente unidimensional.

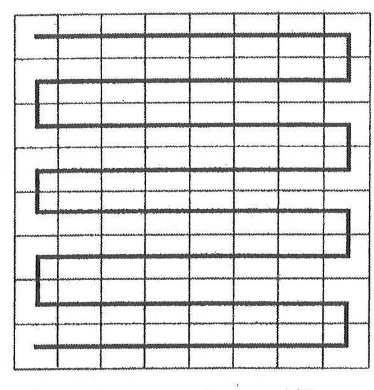
Por tanto, se han ideado una serie de técnicas de búsqueda que persiguen distorsionar al mínimo dominios bidimensionales al trasladarlos al formato unidimensional.

Estas técnicas se basan en la aplicación de una función llamada *índice mosaico* (*tile index)*, que tiene la propiedad de que los puntos que están próximos en el plano euclídeo, estarán próximos también al "camino" que la función dibuja.

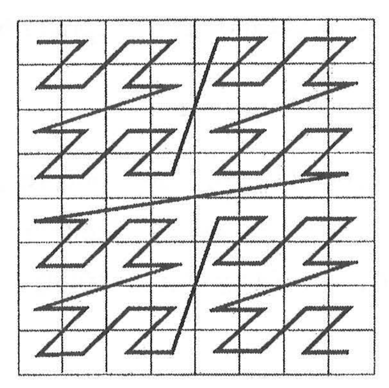
Y en el caso de que se trabaje sobre una retícula discreta, el índice traza un camino por la trama plana, y aunque la propiedad no se verifica exactamente en todos los puntos, los índices mosaico resultan suficientemente eficaces en la mayor parte de los casos en los que se aplican.



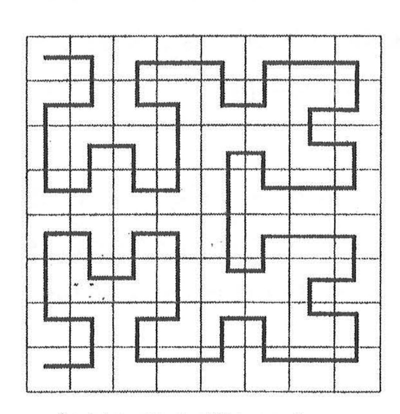
INDICE DE ORDEN POR FILAS



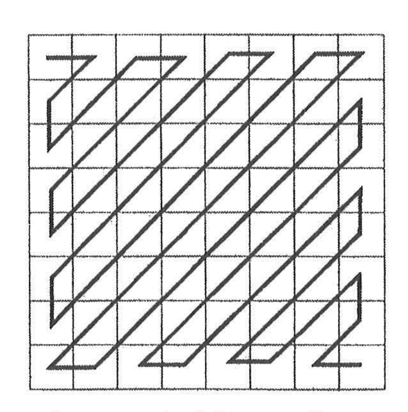
INDICE DE ORDEN PRIMANDO LA FILA



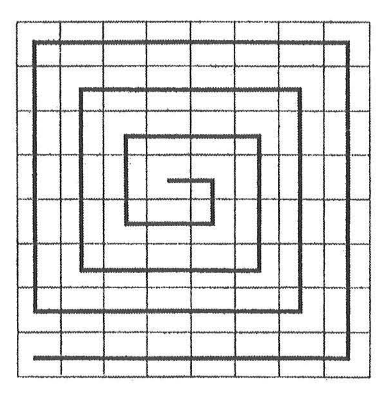
INDICE DE MORTON



INDICE DE PEANO-HILBERT



INDICE DIAGONAL DE CANTOR



INDICE DE ORDEN ESPIRAL

Comentarios bibliográficos

Entre los textos de consulta general sobre las bases de datos, se puede citar el de Elmasri y Navathe (1994), y el introductorio de Date (1995).

Sobre la organización de los archivos, Ferrer i Juliá *et alii* (1997) en el capítulo 1, y Burrough (1986) en la primera parte del capítulo2, presentan el tema en lo esencial y con claridad; y Smith y Barnes (1987) lo abordan con detalle.

Para profundizar en las consultas espaciales, Abel y Mark (1990) presentan las diferentes posibilidades, y Worboys (1995) describe brevemente en el capítulo 6 las diferentes técnicas de búsqueda.

La Parte 5 del libro de Cebrián (1994) profundiza sobre las técnicas concretas de indexación. También se encuentran referencias más generales en los capítulos que dedican a las definiciones de estructuras raster Gutiérrez Puebla y Gould (1994).

Bibliografía citada

ABEL, D.J. y MARK, D.M. (1990): "A comparative analysis of some two-dimensional orderings". *International Journal of Geographical Information Systems* 4: 21-31.

ALDENDERFER, M. y MASCHNER, H.D.G. (Eds.) (1996): *Anthropology, Space, and Geographic Information Systems.* New York / Oxford: Oxford University Press.

ARNAUDIÈRES, J.M. y FRAYSSE, M. (1988): *Cours de Mathématiques - 2. Analyse.* Paris: Dunod Université.

BOSQUE SENDRA, J. (1992): *Sistemas de Información Geográfica.* Madrid: Ediciones Rialp.

BRACKEN, I. y WEBSTER, C. (1989): "Towards a typology of geographical information systems". *International Journal of Geographical Information Systems* 3: 137-152.
(1990): *Information Technology in Geography and Planning. Including Principles of GIS.* London: Routledge.

BURROUGH, P.A. (1986): *Principles of Geographical Information Systems for Land Resources Assessment.* Oxford, Clarendon Press.

CASAS SÁNCHEZ, B, (1993): "Importancia de las nuevas técnicas cartográficas en el conocimiento y evaluación de superficies embalsadas". *Actas del 2° Congreso AESIG.* Madrid: AESIG: 393-406.

CASTRO RÍOS, R. y GARCÍA-ABAD, J.J. (1993): "Confección de la cartografía dinámica de ocupación del suelo con SIG: municipio de Brea de Tajo". *Actas del 2° Congreso AESIG.* Madrid: AESIG: 375-392.

CEBRIÁN, J.A. (1994): *GIS Concepts.*Cáceres: Universidad de Extremadura.

CEBRIÁN, J.A. y MARK, D. (1986): "Modelos topográficos digitales". En *Métodos cuantitativos en Geografía: enseñanza, investigación y planeamiento.* Madrid, A.G.E.: 292-334.

CHIAS NAVARRO, P. (1997): *Los Sistemas de Información Geográfica (I): Introducción y conceptos generales.* Madrid: Publicaciones de la Escuela Técnica Superior de Arquitectura.
(2002): *Aplicación de los Sistemas de Información Geográfica a la redacción de planeamiento considerando las capacidades ambientales del territorio.* Madrid, Instituto Juan de Herrera. Cuadernos de Investigación Urbanística nº 34.

CHÍAS *et alii* (2004): "Nuevos medios para la catalogación y difusión del patrimonio cultural: investigación sobre una metodología innovadora [*Proyecto CICYT FEDER* DPI 2002-04228-C02-02]". En *Actas del X Congreso Internacional de Expresión Gráfica Arquitectónica.* Granada, 6-8 de mayo de 2004). Universidad de Granada, 2004. Pp. 897-905.

DATE, C.J. (1995): *An Introduction to Database Systems.* Reading, MA: Addison-Wesley.

DOUGLAS, D. (1990): "It makes me so CROSS". En PEUQUET, D.J. y MARBLE, D.F. (Eds.) (1990): *Introductory Readings in Geographic Information Systems.* London: Taylor & Francis: 306-307.

ELMASRI, R. y NAVATHE, S.B. (1994): *Fundamentals of Database Systems.* Redwood City, CA: Benjamin / Cummings.

ESRI (1990): *Understanding GIS.* Redlands, CA: ESRI

FERRER, M. *et alii* (1997): *Introducción a los Sistemas de Información Geográfica.* Salamanca: Universidad de Salamanca.

GARCÍA GARCÍA, J. y LÓPEZ PELLICER, M. (1984): *Álgebra lineal y Geometría.* Alcoy: Ed. Marfil.

GARCÍA PALOMO, J.P. y MEDINA SALCEDO, F. (1993): "Estudios atmosféricos desarrollados con SIG tridimensionales". *Actas del 2º Congreso AESIG.* Madrid: AESIG: 415-422.

GATRELL, A.C. (1991): "Concepts of space and geographical data". En MAGUIRE, GOODCHILD y RHIND (Eds.) (1991): *Geographical Information Systems.* Harlow: Longmans. Vol I: 119-134.

GOMBRICH, E.H. (1982): *Arte e ilusión.* Barcelona, G. Gili.

GUTIÉRREZ PUEBLA, J. y GOULD, M. (1994): *SIG: Sistemas de Información Geográfica.* Madrid: Síntesis.

MAGUIRE, D.J.; GOODCHILD, M.F. y RHIND, D.W. (Eds.) (1991): *Geographical Information Systems.* Vol I. Harlow: Longmans.

McHARG, I.L. (1969): *Design with nature.* New York, Doubleday/Natural History Press.

MOREIRA, J.M y GIMÉNEZ AZCÁRATE, F. (1993): "El Sistema de Información Ambiental de Andalucía: SinambA. Una herramienta para el análisis y la gestión del Medio Ambiente". *Actas del 2º Congreso AESIG.* Madrid: AESIG: 491-498.

OPENSHAW, S. (1991): "Developing appropiate spatial analysis methods for GIS". En MAGUIRE, D.J.; GOODCHILD, M.F. y RHIND, D.W. (Eds.) (1991): *Geographical Information Systems.* Harlow: Longmans. Vol I: 389-402.

PUIG ADAM, P. (1986): *Curso de Geometría Métrica.* 2 Vols. Madrid: Euler Ed.

SALAS, J. y CHUVIECO, E. (1993): "Modelización de variables microclimáticas a partir de un Sistema de Información Geográfica". *Actas del 2º Congreso AESIG.* Madrid: AESIG: 423-446.

SMITH, P. y BARNES, G. (1987): *Files and Databases: An Introduction.* Reading, MA: Addison - Wesley.

STEINITZ, C. (1993a): "GIS: A personal historical perspective". *GIS Europe* 5, vol 2: 19-22.
(1993b): "A framework for theory and practice in landscape planning". *GIS Europe* 6, vol 2: 42-45.
(1993c): "The changing face of GIS from 1965-1993". *GIS Europe* 7, vol 2: 38-40.

TOMLIN, C.D. (1990): *Geographic Information Systems and Cartographic Modelling.* Englewood Cliffs, NJ: Prentice Hall.
(1991): "Cartographic Modelling". En MAGUIRE, GOODCHILD y RHIND (Eds.) (1991): *Geographical Information Systems.* Harlow: Longmans. Vol I: 361-374.

TUORI, M. y MOON, G. (1984): "A Topographic Map Conceptual Data Model". *Proceedings of the First International Symposium on Spatial Data Handling.* Zürich. Vol. 1: 28-37.

VELAO, F. (1997): *Los Ayuntamientos, la informatización y el agua milagrosa.* (Conferencia inédita celebrada en la ETSAM el 12-XII-1997).

VILLA, A. de la (1994): *Problemas de Álgebra, con esquemas teóricos.* Madrid: I.C.A.I.

WAUGH,T.C. y HEALEY, R.G. (1986): "The Geoview Design: A Relational database Approach to Geographical data Handling". *Proceedings of the Second International Symposium on Spatial Data Handling.* Seattle (Washington): 193-212. Reproducido también en (1987): *International Journal of GIS* 1: 101-118.

WORBOYS, M.F. (1995): *GIS. A Coomputing Perspective.* London / Bristol PA: Taylor & Francis.

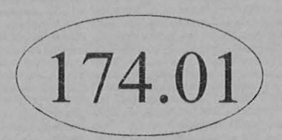
WORBOYS, M.F., HEARNSHAW, H.M. y MAGUIRE, D.J. (1990): "Object oriented data modelling for spatial databases". *International Journal of Geographical Information Systems* 4: 369-383.

NOTAS

NOTAS

NOTAS

CUADERNO



CATÁLOGO Y PEDIDOS EN

http://www.aq.upm.es/of/jherrera

info@mairea-libros.com

I.S.B.N. 84-9728-117-9



84-9728-117-9